

# Spectral Concentration and Greedy $k$ -Clustering

Tamal K. Dey <sup>\*</sup>      Pan Peng<sup>†</sup>      Alfred Rossi <sup>‡</sup>      Anastasios Sidiropoulos <sup>§</sup>

June 25, 2015

## Abstract

A popular graph clustering method is to consider the embedding of an input graph into  $\mathbb{R}^k$  induced by the first  $k$  eigenvectors of its Laplacian, and to partition the graph via geometric manipulations on the resulting metric space. Despite the practical success of this methodology, there is limited understanding of several heuristics that follow this framework. We provide theoretical justification for one such natural and computationally efficient variant.

Our result can be summarized as follows. A partition of a graph is called *strong* if each cluster has small external conductance, and large internal conductance. We present a simple greedy spectral clustering algorithm which returns a partition that is provably close to a suitably strong partition, provided that such a partition exists. A recent result shows that strong partitions exist for graphs with a sufficiently large spectral gap between the  $k$ -th and  $(k + 1)$ -th eigenvalues. Taking this together with our main theorem gives a spectral algorithm which finds a partition close to a strong one for graphs with large enough spectral gap. We also show how this simple greedy algorithm can be implemented in near-linear time for any fixed  $k$  and error guarantee. Finally, we evaluate our algorithm on some real-world and synthetic inputs.

## 1 Introduction

Spectral clustering of graphs is a fundamental technique in data analysis that has enjoyed broad practical usage because of its efficacy and simplicity. The technique maps the vertex set of a graph into a Euclidean space  $\mathbb{R}^k$  where a classical clustering algorithm (such as  $k$ -means,  $k$ -center) is applied to the resulting embedding [VL07]. The coordinates of the vertices in the embedding are computed from  $k$  eigenvectors of a matrix associated with the graph. The exact choice of matrix depends on the specific application but is typically some weighted variant of  $I - A$ , for a graph with adjacency matrix  $A$ .

Despite widespread usage, theoretical understanding of the technique remains limited. Although the case for  $k = 2$  (two clusters) is well understood, the case of general  $k$  is not yet settled and a growing body of work seeks to address the practical success of spectral clustering methods [BXKS11, BLR10, Kel06, NJW01, ST96, VL07]. Recently, the authors in [PSZ14] have shed some light in this direction by showing that algorithms which approximate  $k$ -means applied to the spectral embedding yield  $k$ -clusterings with provable quality guarantees. However, successful application of  $k$ -means is not without practical difficulty, as it is NP-hard and notoriously sensitive to the initial choice of  $k$  centers. Further,  $k$ -means is NP-hard to even approximate to within some constant factor [ACKS15]. Therefore, there is a need to devise a simple fast algorithm that overcomes these bottlenecks, but still achieves a provably good clustering.

---

<sup>\*</sup>Dept. of Computer Science and Engineering, and Dept. of Mathematics, The Ohio State University. Columbus, OH, 43201. [tamaldey@cse.ohio-state.edu](mailto:tamaldey@cse.ohio-state.edu)

<sup>†</sup>Department of Computer Science, TU Dortmund; State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences. [pan.peng@tu-dortmund.de](mailto:pan.peng@tu-dortmund.de)

<sup>‡</sup>Dept. of Computer Science and Engineering, The Ohio State University. Columbus, OH, 43201. [rossi.49@osu.edu](mailto:rossi.49@osu.edu)

<sup>§</sup>Dept. of Computer Science and Engineering, and Dept. of Mathematics, The Ohio State University. Columbus, OH, 43201. [sidiropoulos.1@osu.edu](mailto:sidiropoulos.1@osu.edu)

In this paper we present a simple greedy spectral clustering algorithm which is guaranteed to return a high quality partition, provided that one of sufficient quality *exists*. The resulting partition is close in symmetric difference to the high quality one. Our results can be viewed as providing further theoretical justification for popular clustering algorithms such as in [BXKS11] and [NJW01].

**Measuring partition quality.** Intuitively, a good  $k$ -clustering of a graph is one where there are few edges between vertices residing in different clusters and where each cluster is well-connected as an induced subgraph. Such a qualitative definition of clusters can be appropriately characterized by vertex sets with small external conductance and large internal conductance.

For a subset  $S \subset V$ , the *external conductance* and *internal conductance* are defined to be

$$\varphi_{\text{out}}(S; G) := \frac{|E(S, V(G) \setminus S)|}{\text{vol}(S)}, \quad \varphi_{\text{in}}(S) := \min_{S' \subseteq S, \text{vol}(S') \leq \frac{\text{vol}(S)}{2}} \varphi_{\text{out}}(S'; G[S])$$

respectively, where  $\text{vol}(S) = \sum_{v \in S} \deg(v)$ ,  $E(X, Y)$  denotes the set of edges between  $X$  and  $Y$ , and  $G[S]$  denotes the subgraph of  $G$  induced on  $S$ . When  $G$  is understood from context we sometimes write  $\varphi_{\text{out}}(S)$  in place of  $\varphi_{\text{out}}(S; G)$ .

We define a  $k$ -partition of a graph  $G$  to be a partition  $\mathcal{A} = \{A_1, \dots, A_k\}$  of  $V(G)$  into  $k$  disjoint subsets. We say that  $\mathcal{A}$  is  $(\alpha_{\text{in}}, \alpha_{\text{out}})$ -strong, for some  $\alpha_{\text{in}}, \alpha_{\text{out}} \geq 0$ , if for all  $i \in \{1, \dots, k\}$ , we have

$$\varphi_{\text{in}}(A_i) \geq \alpha_{\text{in}} \quad \text{and} \quad \varphi_{\text{out}}(A_i) \leq \alpha_{\text{out}}.$$

Thus a high quality partition is one where  $\alpha_{\text{in}}$  is large and  $\alpha_{\text{out}}$  is small.

**Our contribution.** We present a simple and fast spectral algorithm which computes a partition provably close to *any*  $(\alpha_{\text{in}}, \alpha_{\text{out}})$ -strong  $k$ -partition if  $\alpha_{\text{in}}$  is large enough (see Theorem 2.1). We emphasize the fact that the algorithm's output approximates any good existing clustering in the input graph. The algorithm consists of a simple greedy clustering procedure performed on the embedding into  $\mathbb{R}^k$  induced by the first  $k$  eigenvectors.

**Related work.** The discrete version of Cheeger's inequality asserts that a graph admits a bipartition into two sets of small external conductance if and only if the second eigenvalue is small [Alo86, AM85, Che70, Mih89, SJ89]. In fact, such a bipartition can be efficiently computed via a simple algorithm that examines the second eigenvector. Generalizations of Cheeger's inequality have been obtained by Lee, Oveis Gharan, and Trevisan [LOT12], and Louis *et al.* [LRTV12]. They showed that spectral algorithms can be used to find  $k$  disjoint subsets, each with small external conductance, provided that the  $k$ -th eigenvalue is small. An improved version of Cheeger's inequality has been obtained by Kwok *et al.* [KLL<sup>+</sup>13] for graphs with large  $k$ -th eigenvalue.

Even though the clusters given by the above spectral partitioning methods have small external conductance, they are not guaranteed to have large internal conductance. In other words, for a resulting cluster  $C$ , the induced graph  $G[C]$  might admit further partitioning into sub-clusters of small conductance. Kannan, Vempala and Vetta proposed quantifying the quality of a partition by measuring the internal conductance of clusters [KVV04].

One may wonder under what conditions a graph admits a partition which provides guarantees on both internal and external conductance. Oveis Gharan and Trevisan, improving on a result of Tanaka [Tan11], showed that graphs which have a sufficiently large spectral gap between the  $k$ -th and  $(k+1)$ -th eigenvalues of its Laplacian admit a  $k$ -clustering which is  $(\alpha \cdot \lambda_{k+1}/k, \beta \cdot k^3 \sqrt{\lambda_k})$ -strong [OT14], for universal constants  $\alpha$ , and  $\beta$ .

Subsequent to the original ArXiv submission [DRS14] of this paper, Peng, Sun, and Zanetti [PSZ14] have used a similar approach to obtain bounds relating the spectral gap to the quality of  $k$ -means clustering performed on the eigenspace. We note that their result differs from ours in that they analyze a different clustering algorithm.

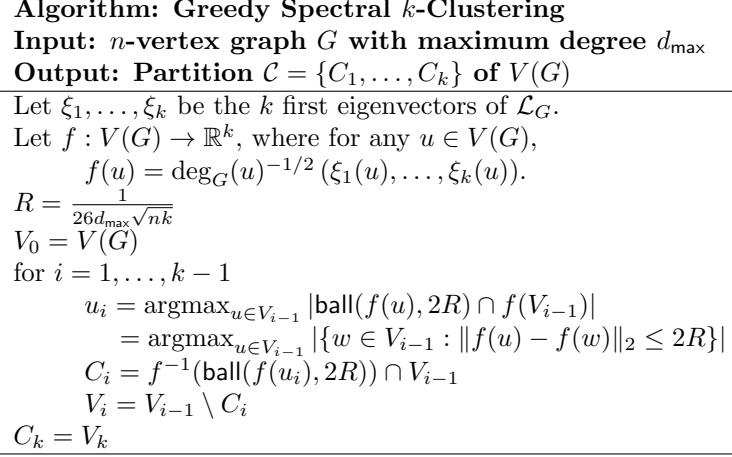


Figure 1: The greedy spectral  $k$ -clustering algorithm.

## 2 Greedy $k$ -Clustering

Let  $G$  be an undirected  $n$ -vertex graph, and let  $\mathcal{L}_G = I - D^{-1/2}AD^{-1/2}$  be its normalized Laplacian, where  $A$  is the adjacency matrix of  $G$  and  $D$  is a diagonal matrix with the entries  $D_{ii}$  equal to the degree of the  $i$ th vertex. Let  $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  be the eigenvalues of  $\mathcal{L}_G$ , and  $\xi_1, \xi_2, \dots, \xi_n \in \mathbb{R}^n$  a corresponding collection of orthonormal left eigenvectors<sup>1</sup>. In this paper we consider a simple geometric clustering operation on the embedding  $f(u)$  which carries a vertex  $u$  to a point given by a rescaling of the first  $k$  eigenvectors of  $\mathcal{L}_G$ ,

$$f(u) = \deg_G(u)^{-1/2} (\xi_1(u), \dots, \xi_k(u)),$$

where  $\deg_G(u)$  denotes the degree of  $u$  in  $G$ . For any  $U \subseteq V(G)$ , let  $f(U) = \{f(u) : u \in U\}$ .

The algorithm iteratively chooses a vertex that has maximum number of vertices within distance  $R$  in  $\mathbb{R}^k$  where  $R$  is computed from  $n, k$ , and  $d_{\max}$ . We treat every such chosen vertex as the “center” of a cluster. For successive iterations, all vertices in previously chosen clusters are discarded. We formally describe the process below.

Inductively define a partition  $\mathcal{C} = \{C_1, \dots, C_k\}$  of  $V(G)$  that uses an auxiliary sequence  $V(G) = V_0 \supseteq V_1 \supseteq \dots \supseteq V_k$ . For any  $i \in \{1, \dots, k-1\}$  and a chosen  $R > 0$ , we proceed as follows. For any  $u \in V_{i-1}$ , let

$$N_i(u) = f^{-1}(\operatorname{ball}(f(u), 2R)) \cap V_{i-1} = \{w \in V_{i-1} : \|f(u) - f(w)\|_2 \leq 2R\}$$

and let  $u_i \in V_{i-1}$  be a vertex maximizing  $|N_i(u)|$ . We set  $C_i = N_i(u_i)$ , and  $V_i = V_{i-1} \setminus C_i$ . Finally, we set  $C_k = V_k$ . This completes the definition of the partition  $\mathcal{C} = \{C_1, \dots, C_k\}$ . The algorithm is summarized in Figure 1. In Section 5, we show how the algorithm can be implemented in time  $O(\varepsilon^{-1}k^2n \log n)$  via random sampling for any error parameter  $\varepsilon > 0$ .

**A distance on  $k$ -partitions.** For two sets  $Y, Z$ , their symmetric difference is given by  $Y \triangle Z = (Y \setminus Z) \cup (Z \setminus Y)$ . Let  $X$  be a finite set,  $k \geq 1$ , and let  $\mathcal{A} = \{A_1, \dots, A_k\}$ ,  $\mathcal{A}' = \{A'_1, \dots, A'_k\}$  be collections of disjoint subsets of  $X$ . Then, we define a distance function between  $\mathcal{A}, \mathcal{A}'$ , by

$$|\mathcal{A} \triangle \mathcal{A}'| = \min_{\sigma} \sum_{i=1}^k |A_i \triangle A'_{\sigma(i)}|,$$

where  $\sigma$  ranges over all permutations  $\sigma$  on  $\{1, \dots, k\}$ .

<sup>1</sup>All vectors in this paper are referred to row vectors.

## 2.1 Main Theorem

**Theorem 2.1** (Spectral partitioning via greedy  $k$ -clustering). *Let  $G$  be an  $n$ -vertex graph with maximum degree  $d_{\max}$ . Let  $k \geq 1$ , and let  $\mathcal{A} = \{A_1, \dots, A_k\}$  be any  $(\alpha_{\text{in}}, \alpha_{\text{out}})$ -strong  $k$ -partition of  $V(G)$  with  $\alpha_{\text{in}} \geq 9(kd_{\max})^{3/2}\sqrt{\lambda_k}$ . Then, on input  $G$ , the algorithm in Figure 1 outputs a partition  $\mathcal{C}$  such that*

$$|\mathcal{A} \triangle \mathcal{C}| = O\left(\frac{\lambda_k \cdot d_{\max}^3 k^4 \cdot n}{\alpha_{\text{in}}^2}\right).$$

We remark that while  $\alpha_{\text{out}}$  does not appear explicitly in the error term in Theorem 2.1, it does implicitly bound the error through a higher order Cheeger inequality [LOT12]. In particular,  $\lambda_k \leq 2\alpha_{\text{out}}$  and thus when  $\alpha_{\text{out}}/\alpha_{\text{in}}^2$  is small there is strong agreement between  $\mathcal{A}$  and  $\mathcal{C}$ .

**Application of main theorem.** Oveis Gharan and Trevisan [OT14] (see also [Tan11]) showed that, if the gap between  $\lambda_k$  and  $\lambda_{k+1}$  is large enough, then there exists a partition into  $k$  clusters, each having small external conductance and large internal conductance.

**Theorem 2.2** ([OT14]). *There exist universal constants  $c > 0$ ,  $\alpha > 0$ , and  $\beta > 0$ , such that for any graph  $G$  with  $\lambda_{k+1} > ck^2\sqrt{\lambda_k}$ , there is a  $(\alpha \cdot \lambda_{k+1}/k, \beta \cdot k^3\sqrt{\lambda_k})$ -strong  $k$ -partition of  $G$ .*

The same paper [OT14] also shows how to compute a partition with slightly worse quantitative guarantees, using an iterative combinatorial algorithm with polynomial running time. More specifically, given a graph  $G$  with  $\lambda_{k+1} > 0$  for any  $k \geq 1$ , their algorithm outputs an  $\ell$ -partition that is  $(\Omega(\lambda_{k+1}^2/k^4), O(k^6\sqrt{\lambda_k}))$ -strong, for some  $1 \leq \ell < k+1$ .

Now we present the following direct corollary of our main theorem.

**Corollary 2.3.** *Let  $G$  be an  $n$ -vertex graph with maximum degree  $d_{\max}$ . Let  $k \geq 1$ , and  $\lambda_{k+1} > \tau k^2\sqrt{\lambda_k}$ , where  $\tau \geq \max\{c, 9d_{\max}^{3/2}k^{1/2}/\alpha\}$  and  $\alpha, c$  are the constants given in Theorem 2.2. Let  $\mathcal{A}$  be the  $k$ -partition of  $G$  guaranteed by Theorem 2.2. Then, on input  $G$ , the algorithm in Figure 1 outputs a partition  $\mathcal{C}$  such that*

$$|\mathcal{A} \triangle \mathcal{C}| = O\left(\frac{d_{\max}^3 k^2 n}{\tau^2}\right).$$

## 3 Spectral Concentration

In this section, we prove that  $\xi_i D^{-1/2}$  for any of the first  $k$  eigenvectors  $\xi_i$  is close (with respect to the  $\ell_2$  norm) to some vector  $\tilde{\mathbf{x}}_i$ , such that  $\tilde{\mathbf{x}}_i$  is constant on each cluster.

**Lemma 3.1.** *Let  $G$  be a graph of maximum degree  $d_{\max}$  with  $\xi_1, \dots, \xi_k \in \mathbb{R}^k$  denoting the first  $k$  eigenvectors of  $\mathcal{L}_G$ . Let  $\mathcal{A} = \{A_1, \dots, A_k\}$  be any  $(\alpha_{\text{in}}, \alpha_{\text{out}})$ -strong  $k$ -partition of  $V(G)$ . For any  $i \in \{1, \dots, k\}$ , if  $\mathbf{x}_i = \xi_i D^{-1/2}$ , then there exists  $\tilde{\mathbf{x}}_i \in \mathbb{R}^n$ , such that,*

$$(i) \quad \|\mathbf{x}_i - \tilde{\mathbf{x}}_i\|_2^2 \leq \frac{2k\lambda_k \cdot d_{\max}^3}{\alpha_{\text{in}}^2}, \text{ and}$$

$$(ii) \quad \tilde{\mathbf{x}}_i \text{ is constant on the clusters of } \mathcal{A}, \text{ i.e. for any } A \in \mathcal{A}, u, v \in A, \text{ we have } \tilde{\mathbf{x}}_i(u) = \tilde{\mathbf{x}}_i(v).$$

Before laying out the proof, we provide some explanation of the statement of the theorem. First, note that the  $\ell_2^2$ -distance between  $\mathbf{x}_i$  and its uniform approximation  $\tilde{\mathbf{x}}_i$  depends linearly on the ratio  $\lambda_k/\alpha_{\text{in}}^2$ , which, as noted above, is bounded from above by  $2\alpha_{\text{out}}/\alpha_{\text{in}}^2$ . Second, the partition-wise uniform vector  $\tilde{\mathbf{x}}_i$  which minimizes the left hand side of (i) is constructed by taking the mean values of  $\mathbf{x}_i$  on each partition. This, together with the bound in (i), means that  $\mathbf{x}_i$  assumes values in each partition close to their mean. In summary, if there is a sufficiently large gap between the external conductance  $\alpha_{\text{out}}$  and internal conductance  $\alpha_{\text{in}}$  of the clusters, the values taken by each vector  $\mathbf{x}_i$  have  $k$  prominent modes over  $k$  partitions.

We need the following result that is a corollary of a lemma in [CPS15] to prove Lemma 3.1. (For completeness, a proof of Lemma 3.2 is included in the appendix.)

**Lemma 3.2.** *Let  $G = (V, E)$  be any undirected graph and let  $C \subseteq V$  be any subset with  $\varphi(G[C]) \geq \varphi_{\text{in}} > 0$ . Then for every  $i$ ,  $1 \leq i \leq k$ ,  $\mathbf{x}_i = \xi_i D^{-1/2}$ , the following holds:*

$$\sum_{u,v \in C} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 \leq \frac{4\lambda_k \cdot \text{vol}(C)}{\varphi_{\text{in}}^2}.$$

*Proof of Lemma 3.1.* Let  $1 \leq i \leq k$ , and  $1 \leq j \leq k$ . By precondition of the lemma,  $\mathcal{A}$  is an  $(\alpha_{\text{in}}, \alpha_{\text{out}})$ -strong  $k$ -partition. Now we apply  $C = A_j$  and  $\varphi_{\text{in}} = \alpha_{\text{in}}$  in Lemma 3.2 to get

$$\sum_{u,v \in A_j} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 \leq \frac{4\lambda_k \cdot \text{vol}(A_j)}{\alpha_{\text{in}}^2} \leq \frac{4\lambda_k \cdot d_{\text{max}} \cdot |A_j|}{\alpha_{\text{in}}^2},$$

where the second inequality follows from our assumption that the maximum degree is  $d_{\text{max}}$ . On the other hand, by the definition of  $\tilde{\mathbf{x}}_i$ , we have

$$\frac{1}{|A_j|} \sum_{u,v \in A_j} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 = 2 \sum_{u \in A_j} (\mathbf{x}_i(u) - \tilde{\mathbf{x}}_i(u))^2.$$

Therefore,

$$\|\mathbf{x}_i - \tilde{\mathbf{x}}_i\|_2^2 = \sum_{j=1}^k \sum_{u \in A_j} (\mathbf{x}_i(u) - \tilde{\mathbf{x}}_i(u))^2 = \frac{1}{2} \sum_{j=1}^k \frac{1}{|A_j|} \sum_{u,v \in A_j} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 \leq \frac{2k\lambda_k \cdot d_{\text{max}}}{\alpha_{\text{in}}^2}.$$

where the penultimate inequality follows from our assumption that  $\lambda_{k+1} \geq \tau k^2 \sqrt{\lambda_k}$ . This completes the proof of the lemma.  $\square$

## 4 From Spectral Concentration to Spectral Clustering

In this section we prove Theorem 2.1. We begin by showing that in the embedding induced by the  $k$  vectors  $\xi_1 D^{-1/2}, \dots, \xi_k D^{-1/2}$ , most of the clusters in any given  $k$ -partition with high quality are concentrated around center points in  $\mathbb{R}^k$  which are sufficiently far apart from each other.

**Lemma 4.1.** *Let  $G$  be an  $n$ -vertex graph of maximum degree  $d_{\text{max}}$ . Let  $\mathcal{A} = \{A_1, \dots, A_k\}$  be an  $(\alpha_{\text{in}}, \alpha_{\text{out}})$ -strong  $k$ -partition of  $V(G)$  with  $\alpha_{\text{in}} \geq 9(kd_{\text{max}})^{3/2} \sqrt{\lambda_k}$ . Let  $\xi_1, \dots, \xi_n$  be the eigenvectors of  $\mathcal{L}_G$ , and let  $f : V(G) \rightarrow \mathbb{R}^k$  be the spectral embedding of  $G$  such that for any  $u \in V(G)$ ,  $f(u) = (\xi_1(u), \dots, \xi_k(u))(\deg_G(u))^{-1/2}$ . Let  $R = \frac{1}{26d_{\text{max}}\sqrt{nk}}$ . Then there exists a set  $\mathcal{A}' = \{A'_1, \dots, A'_k\}$  of  $k$  disjoint subsets of  $G$ , and  $\mathbf{p}_1, \dots, \mathbf{p}_k \in \mathbb{R}^k$ , such that the following conditions are satisfied:*

- (i)  $|\mathcal{A} \triangle \mathcal{A}'| = O\left(\frac{\lambda_k \cdot d_{\text{max}}^3 k^3 \cdot n}{\alpha_{\text{in}}^2}\right)$ .
- (ii) For  $1 \leq i \leq k$ ,  $A'_i = \{u \in V : \|f(u) - \mathbf{p}_i\|_2 \leq R\}$ .
- (iii) For  $1 \leq i < j \leq k$ ,  $\|\mathbf{p}_i - \mathbf{p}_j\|_2 > 6R$ .

To prove the item (iii) in the statement of the lemma, we need the following facts. For any symmetric matrix  $X$ , let  $\eta_i(X)$  denote the  $i$ th largest eigenvalue of  $X$ . For any matrix  $Y$ , let  $Y_{\text{row}(i)}$  denote the  $i$ th row vector of  $Y$ .

**Fact 4.2.** *For any two  $p \times p$  symmetric matrices  $X, Y$ , if  $\max_{i \leq p} \|X_{\text{row}(i)} - Y_{\text{row}(i)}\|_2 \leq \delta$ , then for any  $i \leq k$ ,  $|\eta_i(X) - \eta_i(Y)| \leq \sqrt{p}\delta$ .*

*Proof.* Since  $\max_{i \leq p} \|X_{\text{row}(i)} - Y_{\text{row}(i)}\|_2 \leq \delta$ , then the Frobenius norm  $\|X - Y\|_F$  of  $X - Y$  is at most  $\sqrt{p}\delta$ , and therefore, the induced 2-norm  $\|X - Y\|_2$  of  $X - Y$  is at most  $\sqrt{p}\delta$ .

Now by the Weyl's inequality [Tao12], for any  $i$ ,  $|\eta_i(X) - \eta_i(Y)| \leq |\eta_1(X - Y)| \leq \|X - Y\|_2 \leq \sqrt{p}\delta$ . This completes the proof of the fact.  $\square$

**Fact 4.3.** For any two  $p \times q$  matrices  $X, Y$ , if  $\max_{i \leq p} \|X_{\text{row}(i)}\|_2 \leq \gamma$ , and  $\max_{i \leq p} \|X_{\text{row}(i)} - Y_{\text{row}(i)}\|_2 \leq \delta$ , then  $\max_{i \leq p} \|(X \cdot X^T)_{\text{row}(i)} - (Y \cdot Y^T)_{\text{row}(i)}\|_2 \leq \sqrt{p}(\delta^2 + 2\gamma\delta)$ .

*Proof.* For simplicity, let  $X_i, Y_i$  denote the  $i$ th row of  $X, Y$ , respectively. Note that the  $i, j$ th entry of  $X \cdot X^T$  and  $Y \cdot Y^T$  are  $\langle X_i, X_j \rangle$  and  $\langle Y_i, Y_j \rangle$ , respectively, and that

$$\begin{aligned} |\langle X_i, X_j \rangle - \langle Y_i, Y_j \rangle| &= |\langle X_i, X_j \rangle - \langle Y_i - X_i + X_i, Y_j - X_j + X_j \rangle| \\ &= |\langle X_i, X_j \rangle - \langle Y_i - X_i, Y_j - X_j \rangle - \langle Y_i - X_i, X_j \rangle - \langle X_i, Y_j - X_j \rangle - \langle X_i, X_j \rangle| \\ &\leq |\langle Y_i - X_i, Y_j - X_j \rangle| + |\langle Y_i - X_i, X_j \rangle| + |\langle X_i, Y_j - X_j \rangle| \\ &\leq \|Y_i - X_i\|_2 \cdot \|Y_j - X_j\|_2 + \|Y_i - X_i\|_2 \cdot \|X_j\|_2 + \|X_i\|_2 \cdot \|Y_j - X_j\|_2 \\ &\leq \delta^2 + 2\gamma\delta. \end{aligned}$$

Therefore,  $\max_{i \leq p} \|(X \cdot X^T)_{\text{row}(i)} - (Y \cdot Y^T)_{\text{row}(i)}\|_2 \leq \sqrt{p}(\delta^2 + 2\gamma\delta)$ .  $\square$

*Proof of Lemma 4.1.* Recall that for any  $i$ ,  $1 \leq i \leq k$ ,  $\mathbf{x}_i = \xi_i D^{-1/2}$  and  $\tilde{\mathbf{x}}_i$  denotes the vector that  $\tilde{\mathbf{x}}_i(u) = \frac{1}{|A_j|} \sum_{v \in A_j} \mathbf{x}_i(v)$  if  $u \in A_j$ . Now for each  $1 \leq j \leq k$ , define

$$\mathbf{p}_j := (\tilde{\mathbf{x}}_1(u), \dots, \tilde{\mathbf{x}}_k(u)), \quad \text{for any } u \in A_j.$$

Now we prove Item (iii). Let  $X$  be the  $k \times n$  matrix with  $X_{\text{row}(i)} = \mathbf{x}_i$  for each  $i \leq k$ . Let  $\tilde{X}$  be the  $k \times n$  matrix with  $\tilde{X}_{\text{row}(i)} = \tilde{\mathbf{x}}_i$  for each  $i \leq k$ . Note that for any  $u \in A_j$ , the column vector corresponding to  $u$  of  $X$  is  $\mathbf{p}_j^T$ .

First, we note that all the eigenvalues of  $X \cdot X^T$  are at least  $1/d_{\max}$ . This is true since for any  $v \in \mathbb{R}^k$ ,  $v(X \cdot X^T)v^T = \|vX\|_2^2 \geq \|vFD^{-1/2}\|_2^2 \geq \|vF\|_2^2/d_{\max} = v \cdot v^T/d_{\max}$ , where  $F$  is the  $k \times n$  matrix with  $i$ th row  $\xi_i$  for each  $i \leq k$ , and the last equation follows from the observation that  $F \cdot F^T = I_{k \times k}$ .

Now let  $\zeta = \frac{2k\lambda_k \cdot d_{\max}}{\alpha_{\text{in}}^2}$ . Note that by Lemma 3.1, for each  $i \leq k$ ,  $\|X_{\text{row}(i)}\|_2 = \|\mathbf{x}_i\|_2 \leq 1$  and

$$\|X_{\text{row}(i)} - \tilde{X}_{\text{row}(i)}\|_2 = \|\mathbf{x}_i - \tilde{\mathbf{x}}_i\|_2 \leq \sqrt{\zeta}.$$

By Fact 4.3, 4.2 and the fact that all the eigenvalues of  $X \cdot X^T$  are at least  $\frac{1}{d_{\max}}$ , we know that all eigenvalues of  $\tilde{X} \cdot \tilde{X}^T$  are at least  $\frac{1}{d_{\max}} - k(\zeta + 2\sqrt{\zeta})$ .

Item (iii) of the lemma will then follow from the following claim, and the inequality that  $\sqrt{k}(36R^2n + 12R\sqrt{n}) < \frac{1}{d_{\max}} - k(\zeta + 2\sqrt{\zeta})$  by our choice of parameters.

**Claim 4.4.** If there exists  $i_0, j_0 \leq k$  such that  $\|\mathbf{p}_{i_0} - \mathbf{p}_{j_0}\|_2 \leq 6R$ , then  $\tilde{X} \cdot \tilde{X}^T$  has an eigenvalue at most  $\sqrt{k}(36R^2n + 12R\sqrt{n})$ .

*Proof.* Let  $\bar{X}$  denote the  $k \times n$  matrix obtained from  $\tilde{X}$  by replacing each vector that equals  $\mathbf{p}_{i_0}$  by vector  $\mathbf{p}_{j_0}$ . Note that  $\bar{X} \cdot \bar{X}^T$  is singular, and thus has eigenvalue 0.

Now note that  $\max_{i \leq k} \|\tilde{X}_{\text{row}(i)} - \bar{X}_{\text{row}(i)}\|_2 \leq 6R\sqrt{n}$  since the absolute value of each entry in  $\tilde{X} - \bar{X}$  is at most  $6R$ , and also note that for any  $i \leq k$ ,

$$\|\tilde{X}_{\text{row}(i)}\|_2 = \sqrt{\sum_{j=1}^k |A_j| \left( \frac{\sum_{u \in A_j} \mathbf{x}_i(u)}{|A_j|} \right)^2} \leq \sqrt{\sum_{j=1}^k |A_j| \frac{\sum_{u \in A_j} \mathbf{x}_i^2(u)}{|A_j|}} = \|\tilde{\mathbf{x}}_i\|_2 \leq 1.$$

Then by Fact 4.3,  $\max_{i \leq k} \|(\tilde{X} \cdot \tilde{X}^T)_{\text{row}(i)} - (\bar{X} \cdot \bar{X}^T)_{\text{row}(i)}\|_2 \leq \sqrt{k}(36R^2n + 2 \cdot 6R\sqrt{n})$ . Now by Fact 4.2 and the fact that  $\bar{X} \cdot \bar{X}^T$  has eigenvalue 0, we know that at least one eigenvalue of  $\tilde{X} \cdot \tilde{X}^T$  is at most  $\sqrt{k}(36R^2n + 2 \cdot 6R\sqrt{n})$ .  $\square$

Now for each  $1 \leq j \leq k$ , define  $A'_j := \{u \in V : \|f(u) - \mathbf{p}_j\|_2 \leq R\}$  as required by Item (ii). Let  $\mathcal{A}' = \{A'_1, \dots, A'_k\}$ . By Item (iii),  $A'_1, \dots, A'_k$  are disjoint. Recall that  $\zeta = \frac{2k\lambda_k \cdot d_{\max}}{\alpha_{\text{in}}^2}$ , and by Lemma 3.1,

$$\sum_{i=1}^k \|\mathbf{x}_i - \tilde{\mathbf{x}}_i\|_2^2 \leq k \cdot \zeta.$$

On the other hand, if we let  $A_{\text{bad}} = \{u : u \in A_j, \|f(u) - \mathbf{p}_j\|_2 > R, 1 \leq j \leq k\}$ , then

$$\begin{aligned} \sum_{i=1}^k \|\mathbf{x}_i - \tilde{\mathbf{x}}_i\|_2^2 &= \sum_{j=1}^k \sum_{u \in A_j} \sum_{i=1}^k (\mathbf{x}_i(u) - \tilde{\mathbf{x}}_i(u))^2 = \sum_{j=1}^k \sum_{u \in A_j} \|f(u) - \mathbf{p}_j\|_2^2 \\ &\geq \sum_{u \in A_{\text{bad}}} R^2 = |A_{\text{bad}}| \cdot R^2. \end{aligned}$$

Therefore,  $|A_{\text{bad}}| \leq \frac{k\zeta}{R^2} \leq \frac{1500\lambda_k \cdot d_{\max}^3 k^3 \cdot n}{\alpha_{\text{in}}^2}$ . Item (i) in the statement of the lemma then follows from the observation that  $|\mathcal{A} \triangle \mathcal{A}'| \leq 2 \cdot |A_{\text{bad}}|$ .  $\square$

We are now ready to prove our main theorem.

*Proof of Theorem 2.1.* Let  $\mathcal{A} = \{A_1, \dots, A_k\}$ ,  $\mathcal{A}' = \{A'_1, \dots, A'_k\}$ ,  $f$ ,  $R$ , and  $\mathbf{p}_1, \dots, \mathbf{p}_k$  be as in Lemma 4.1. Let  $\varepsilon = |\mathcal{A} \triangle \mathcal{A}'|/n = O(\frac{\lambda_k \cdot d_{\max}^3 k^3}{\alpha_{\text{in}}^2})$ . Let  $\mathcal{C} = \{C_1, \dots, C_k\}$  be the ordered collection of pairwise disjoint subsets of  $V(G)$  output by the greedy spectral  $k$ -clustering algorithm in Figure 1. The set of vertices not covered by any of the clusters in  $\mathcal{A}'$  plays a special role in our argument which we denote as  $B = V(G) \setminus \left(\bigcup_{i=1}^k A'_i\right)$ . Clearly,  $|B| \leq |\mathcal{A} \triangle \mathcal{A}'| \leq \varepsilon n$ .

We say that a cluster  $A'_i$  is *touched* if the algorithm outputs a cluster  $C_j \in \mathcal{C}$  with  $C_j \cap A'_i \neq \emptyset$ . Since the clusters in  $\mathcal{C}$  cover all vertices in  $V(G)$ , every cluster in  $\mathcal{A}'$  is touched by some cluster in  $\mathcal{C}$ . For a cluster  $A'_i$ , let  $C_{\rho(i)}$  be the cluster in  $\mathcal{C}$  that touches  $A'_i$  for the *first time* in the algorithm. Let  $I$  be a maximal subset of  $\{1, \dots, k\}$  such that the restriction of  $\rho$  on  $I$  is a bijection. Let  $i^* = |I|$ . By permuting the indices of the clusters in  $\mathcal{A}'$ , we may assume w.l.o.g. that  $I = \{1, \dots, i^*\}$ . In particular, if  $i^* = k$ , then  $\rho$  is a bijection between  $\{1, \dots, k\}$  and  $\{1, \dots, k\}$ .

If  $i^* < k$ , we claim that the clusters  $\{A'_i \mid i = i^* + 1, \dots, k\}$  are all mapped to  $C_k$ , that is,  $\rho(i^* + 1) = \dots = \rho(k) = k$ . This is because the cluster  $C_{\rho(i)} \neq C_k$  can intersect at most one cluster in  $\mathcal{A}'$  because every cluster in  $\mathcal{A}'$  is contained inside some ball of radius  $R$ , the distance between any two centers of such balls is more than  $6R$ , and each  $C_{\rho(i)} \neq C_k$  is contained inside some ball of radius  $2R$ .

We further observe that if  $i^* < k$ , then a cluster  $A'_i$  with  $\rho(i) = k$  can have at most  $2\varepsilon n$  vertices. Suppose not. By the above claim that  $\rho(i^* + 1) = \dots = \rho(k) = k$ , we know that there is a cluster  $C_j$  with  $j < k$ , which does not intersect any cluster in  $\mathcal{A}'$  for the first time. Then, it has the only option of intersecting a cluster in  $\mathcal{A}'$  beyond the first time and/or intersect  $B$ . Since  $|A'_i \setminus C_{\rho(i)}| \leq \varepsilon n$  for all  $i$ ,  $C_j$  can have at most  $\varepsilon n + |B| \leq 2\varepsilon n$  vertices. But, the algorithm could have made a better choice by selecting  $A'_i$  while computing  $C_j$  because  $|A'_i| > 2\varepsilon n$  by our assumption. We reach a contradiction.

Next, observe that  $|A'_i \setminus C_{\rho(i)}| \leq \varepsilon n$  for every cluster  $A'_i$ . This is certainly true if  $\rho(i) = k$  because then  $C_k$  contains  $A'_i$  completely. When  $\rho(i) \neq k$ ,  $C_{\rho(i)}$  cannot intersect any other cluster in  $\mathcal{A}'$  and it can get at most  $\varepsilon n$  vertices from  $B$ . If  $A'_i \setminus C_{\rho(i)}$  had more than  $\varepsilon n$  vertices, the algorithm could have made a better choice by taking the entire  $A'_i$  while computing  $C_{\rho(i)}$ . Such a choice can be made by taking  $C_{\rho(i)}$  to be all the yet unclustered points that are inside a ball of radius  $2R$  centered at any point in  $A'_i$ ; since  $A'_i$  is in a ball of radius  $R$ , it follows by the triangle inequality that  $A'_i$  will be contained inside  $C_{\rho(i)}$ .

Since for every  $i \leq i^*$ ,  $|A'_i \setminus C_{\rho(i)}| \leq \varepsilon n$  and  $C_{\rho(i)}$  cannot intersect any cluster in  $\mathcal{A}'$  other than  $A'_i$ , we have  $|A'_i \triangle C_{\rho(i)}| \leq \varepsilon n + |B \cap C_{\rho(i)}|$ . If  $i^* < i \leq k$ ,  $C_k$  contains  $A'_i$  entirely and may intersect other clusters in  $\mathcal{A}'$  for the second time and beyond. Then, we have  $|A'_i \triangle C_k| \leq k\varepsilon n + |B \cap C_k|$ . Let  $T = \{1, \dots, k-1\} \setminus \{\rho(1), \dots, \rho(i^*)\}$  be the set of indices  $j$  such that  $C_j$  does not intersect any cluster in  $\mathcal{A}'$  for

---

**Algorithm: Fast Spectral  $k$ -Clustering**

---

**Input:**  $n$ -vertex graph  $G$  with maximum degree  $d_{\max}$ , and  $\varepsilon > 0$ .

**Output:** Partition  $\mathcal{C} = \{C_1, \dots, C_k\}$  of  $V(G)$

---

Let  $\xi_1, \dots, \xi_k$  be the  $k$  first eigenvectors of  $G$ .

Let  $f : V(G) \rightarrow \mathbb{R}^k$ , where for any  $u \in V(G)$ ,

$$f(u) = \deg_G(u)^{-1/2} (\xi_1(u), \dots, \xi_k(u)).$$

$$R = \frac{1}{26d_{\max}\sqrt{nk}}$$

$$V_0 = V(G)$$

for  $i = 1, \dots, k-1$

Sample uniformly with repetition a subset  $U_{i-1} \subseteq V_{i-1}$ ,  $|U_{i-1}| = \Theta(\varepsilon^{-1} \log n)$ .

$$u_i = \operatorname{argmax}_{u \in U_{i-1}} |\operatorname{ball}(f(u), 2R) \cap f(V_{i-1})|$$

$$= \operatorname{argmax}_{u \in U_{i-1}} |\{w \in V_{i-1} : \|f(u) - f(w)\|_2 \leq 2R\}|$$

$$C_i = f^{-1}(\operatorname{ball}(f(u_i), 2R)) \cap V_{i-1}$$

$$V_i = V_{i-1} \setminus C_i$$

$$C_k = V_k$$

---

Figure 2: A faster spectral  $k$ -clustering algorithm.

the first time. For any  $j$  such that  $j \in T$ ,  $C_j$  can only intersect set  $B$  and/or set  $A_i \setminus C_{\rho(i)}$  for any  $i$  that  $\rho(i) < j$ . This then gives that  $|\bigcup_{j \in T} C_j| \leq \sum_{j \in T} |B \cap C_j| + \sum_{i \leq i^*} |A_i \setminus C_{\rho(i)}|$ .

Using the bijectivity of  $\rho$  on the set  $\{1, \dots, i^*\}$ , we have

$$\begin{aligned} |\mathcal{A}' \triangle \mathcal{C}| &\leq \sum_{i \leq i^*} |A'_i \triangle C_{\rho(i)}| + |C_k \triangle A'_{i^*+1}| + |\bigcup_{j \in T} C_j| + |\bigcup_{i \geq i^*+2} A'_i| \\ &\leq k\varepsilon n + \sum_{i \leq i^*} |B \cap C_{\rho(i)}| + k\varepsilon n + |B \cap C_k| + \sum_{j \in T} |B \cap C_j| + k\varepsilon n + |\bigcup_{i \geq i^*+2} A'_i| \\ &\leq 3k\varepsilon n + |B| + 2k\varepsilon n \\ &\leq 6k\varepsilon n. \end{aligned}$$

The following concludes the proof:

$$|\mathcal{A} \triangle \mathcal{C}| \leq |\mathcal{A} \triangle \mathcal{A}'| + |\mathcal{A}' \triangle \mathcal{C}| < \varepsilon n + |\mathcal{A}' \triangle \mathcal{C}| \leq 7k\varepsilon n = O\left(\frac{\lambda_k \cdot d_{\max}^3 k^4 \cdot n}{\alpha_{\text{in}}^2}\right).$$

□

## 5 Implementation in Practice

The greedy algorithm in Figure 1 iterates over all  $n$  vertices to determine the best center. In practice,  $n$  can be much larger than  $k$ . We next show how to overcome this bottleneck via random sampling.

### 5.1 A Faster Algorithm

In the algorithm from the previous section, in every iteration  $i \in \{1, \dots, k\}$ , we compute the value  $|N_i(u)|$  for all  $u \in V_i$ . We can speed up the algorithm by computing  $|N_i(u)|$  only for a randomly chosen subset of  $V_i$ , of size about  $\Theta(\varepsilon^{-1} \log n)$ , for error parameter  $\varepsilon$ . This results in a faster randomized algorithm that runs in  $O(\varepsilon^{-1} n k^2 \log n)$  time. It is summarized in Figure 2. A statement similar to Theorem 2.1.



**Theorem 5.1.** Let  $\varepsilon > 0$ . Let  $G$  be an  $n$ -vertex graph with maximum degree  $d_{\max}$ . Let  $k \geq 1$ , and let  $\mathcal{A} = \{A_1, \dots, A_k\}$  be any  $(\alpha_{\text{in}}, \alpha_{\text{out}})$ -strong  $k$ -partition of  $V(G)$  with  $\alpha_{\text{in}} \geq 9(kd_{\max})^{3/2}\sqrt{\lambda_k}/\sqrt{\varepsilon}$ . Then, on input  $G$ , with high probability, the algorithm in Figure 2 outputs a partition  $\mathcal{C}$  such that

$$|\mathcal{A} \triangle \mathcal{C}| \leq \varepsilon n,$$

Furthermore, the running time of the algorithm is  $O(\varepsilon^{-1}k^2n \log n)$ .

*Proof sketch.* Since  $\alpha_{\text{in}} \geq 9(kd_{\max})^{3/2}\sqrt{\lambda_k}/\sqrt{\varepsilon}$ , then we can apply Lemma 4.1 to find a collection  $\mathcal{A}' = \{A'_1, \dots, A'_k\}$  of pairwise disjoint subsets of  $V(G)$ , such that  $|\mathcal{A} \triangle \mathcal{A}'|/n \leq \varepsilon$ . Similar to the proof of Theorem 2.1, we define a function  $\rho : \{1, \dots, k\} \rightarrow \{1, \dots, k\}$  that maps clusters in  $\mathcal{A}'$  to clusters in  $\mathcal{C}$ , where  $\mathcal{C} = \{C_1, \dots, C_k\}$  are the ordered collection of pairwise disjoint subsets of  $V(G)$  output by the greedy spectral  $k$ -clustering algorithm in Figure 2. A key observation now is that for any  $i \leq k$ , a vertex from a cluster  $A'_i$  of size at least  $2\varepsilon n$  will be sampled out. This implies that with high probability the following holds: for each  $i \in \{1, \dots, k\}$ , if  $|A'_i| \geq 2\varepsilon n$ , then  $\rho(i) \neq k$ , since the vertices in the set  $U_i$  of the  $i$ th iteration of the algorithm are sampled uniformly at random. The rest of the argument for the correctness of the algorithm is identical to the proof of Theorem 2.1. For the running time of the algorithm, note that in each iteration, we need to sample  $O(\varepsilon^{-1} \log n)$  vertices, and for each sampled vertex  $u$ , we need to compute  $|N_i(u)|$ , which takes  $O(nk)$  time. Since there are  $k$  iterations, the total running time of the algorithm is thus  $O(\varepsilon^{-1}nk^2 \log n)$ .  $\square$

In practice it is common to work with fixed  $k$ . We note that the fast randomized algorithm runs in near-linear time for  $k = O(\text{polylog}(n))$ , provided that the user specifies  $\varepsilon = \Omega(1/\text{polylog}(n))$ .

## 5.2 Experimental Evaluation

Results from our greedy  $k$ -clustering implementation are shown in Figures 3, 4. Cluster assignments for graphs are shown as colored nodes. In the case where the graph comes from a triangulated surface, we have extended the coloring to a small surface patch in the vicinity of the node. Each experiment includes a plot of the eigenvalues of the normalized Laplacian. A small rectangle on each plot highlights the corresponding spectral gap between  $k$  and  $k + 1$ .

**Multiple spectral gaps.** Recall that graphs which have a sufficiently large spectral gap between the  $k$ -th and  $(k + 1)$ -th eigenvalues admit a strong clustering [OT14]. Figure 3 shows the result of our algorithm on a graph with two prominent spectral gaps,  $k = 2$  (left) and  $k = 5$  (right).

This graph is sampled from the following generative model. Let  $C_1, \dots, C_5$  be disjoint vertex sets of equal size, depicted as circles. Every edge with both endpoints in the same  $C_i$  appears with probability  $p_1$ , every edge between  $C_1 \cup C_2$  and  $C_3 \cup C_4 \cup C_5$  appears with probability  $p_3$ , and every other edge appears with probability  $p_2$ , for some  $p_1 \gg p_2 \gg p_3$ . The resulting graph admits a strong  $k$ -partition, for any  $k \in \{2, 5\}$ . This fact is reflected in the output of our algorithm.

We remark that to achieve a sufficiently large gap at  $k = 5$  many intra-circle edges are necessary, which makes the resulting figures too dense. To make the plots readable, we have displayed only a subsampling of these edges.

**Comparison with  $k$ -means.** In Figure 4 we compare the greedy approach against  $k$ -means on the spectral embedding, which was computed using SciPy’s `kmeans2` function.

The greedy approach proves favorable except perhaps in *R15* where the results are comparable. Interestingly, some of the  $k$ -means labels on the human model are disconnected on the manifold. This is due to the selection of a center point between protrusions in the embedding.

The poor performance of  $k$ -means might be due to its sensitivity to the initial seed placements, which were made automatically by the default behavior of `kmeans2`. We further remark that our comparison is intended only as a baseline against standard  $k$ -means on the embedding, and was applied directly to the embedding without any preprocessing that could be beneficial to  $k$ -means.

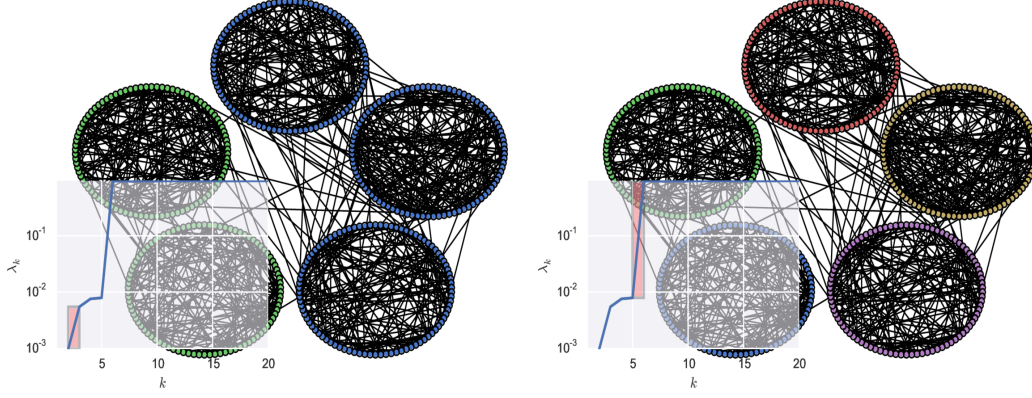


Figure 3: Clustering a graph with multiple spectral gaps.

The spectral gap in the above examples is generally smaller than the requirement in our Theorems. Despite this, our spectral clustering algorithm seems to produce meaningful results even in such examples. This suggests that stronger theoretical guarantees might be obtainable. We believe this is an interesting research direction.

## References

- [ACKS15] Pranjali Awasthi, Moses Charikar, Ravishankar Krishnaswamy, and Ali Kemal Sinop. The hardness of approximation of euclidean k-means. *CoRR*, abs/1502.03316, 2015.
- [Alo86] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.
- [AM85] Noga Alon and V. D. Milman.  $\lambda_1$ , isoperimetric inequalities for graphs, and superconcentrators. *J. Comb. Theory, Ser. B*, 38(1):73–88, 1985.
- [BLR10] Punyashloka Biswal, James R. Lee, and Satish Rao. Eigenvalue bounds, spectral partitioning, and metrical deformations via flows. *J. ACM*, 57(3), 2010.
- [BXKS11] Sivaraman Balakrishnan, Min Xu, Akshay Krishnamurthy, and Aarti Singh. Noise thresholds for spectral clustering. In *NIPS*, pages 954–962, 2011.
- [Che70] Jeff Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in Analysis*, pages 195–199. Princeton Univ. Press, Princeton, NJ, 1970.
- [Chu97] Fan R.K. Chung. *Spectral Graph Theory*. Number no. 92 in CBMS Regional Conference Series. Conference Board of the Mathematical Sciences, 1997.
- [CPS15] Artur Czumaj, Pan Peng, and Christian Sohler. Testing cluster structure of graphs. In *STOC*, 2015.
- [DRS14] Tamal K. Dey, Alfred Rossi, and Anastasios Sidiropoulos. Spectral concentration, robust k-center, and simple clustering. *CoRR*, abs/1404.1008, 2014.
- [Kel06] Jonathan A. Kelner. Spectral partitioning, eigenvalue bounds, and circle packings for graphs of bounded genus. *SIAM J. Comput.*, 35(4):882–902, 2006.
- [KLL<sup>+</sup>13] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, and Luca Trevisan. Improved Cheeger’s inequality: analysis of spectral partitioning algorithms through higher order spectral gap. In *STOC*, pages 11–20, 2013.

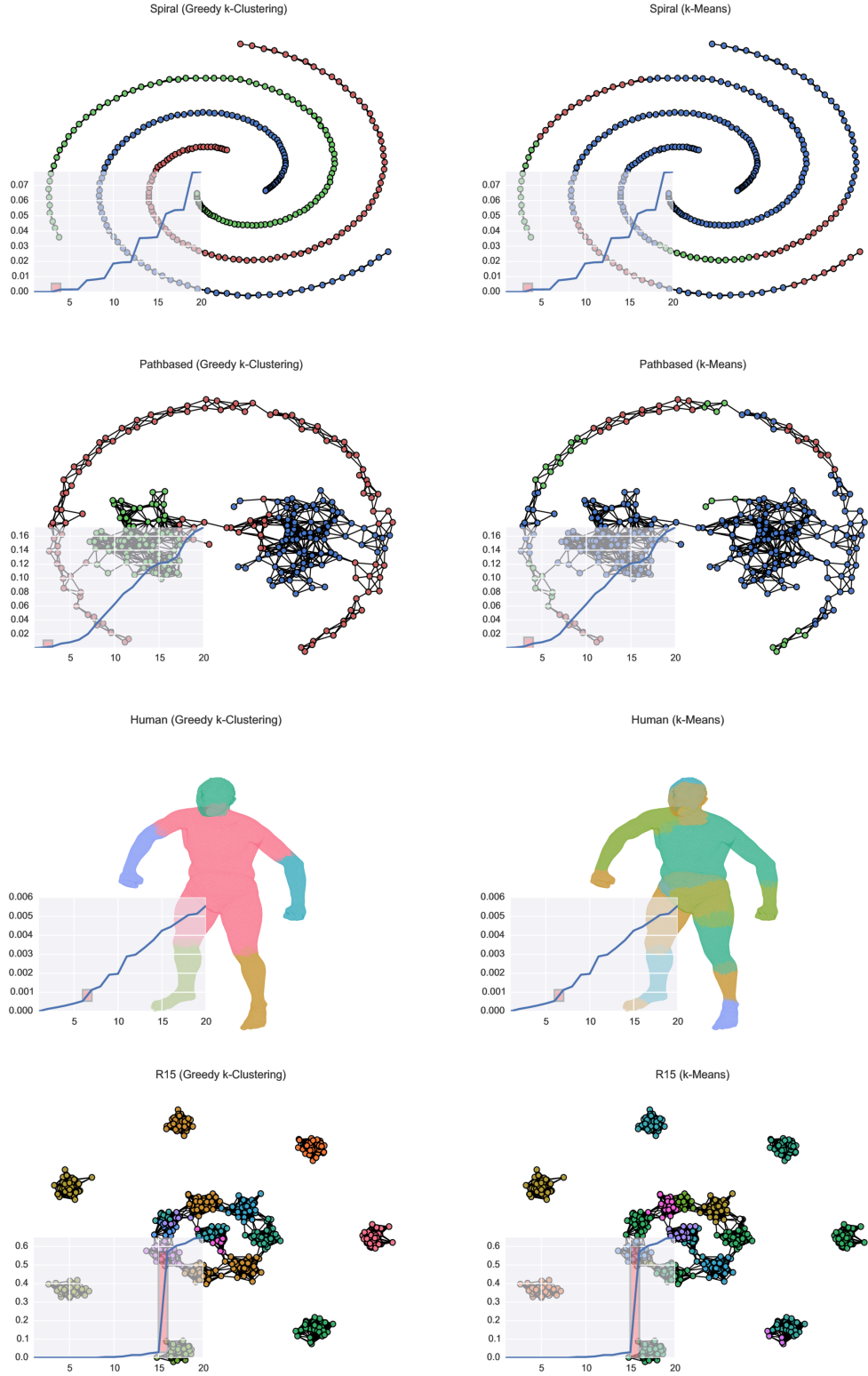


Figure 4: Comparison to  $k$ -means.

- [KVV04] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *J. ACM*, 51(3):497–515, 2004.
- [LOT12] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan. Multi-way spectral partitioning and higher-order cheeger inequalities. In *STOC*, pages 1117–1130, 2012.
- [LRTV12] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh Vempala. Many sparse cuts via higher eigenvalues. In *STOC*, pages 1131–1140, 2012.
- [Mih89] Milena Mihail. Conductance and convergence of markov chains-a combinatorial treatment of expanders. In *FOCS*, pages 526–531, 1989.
- [NJW01] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *NIPS*, pages 849–856, 2001.
- [OT14] Shayan Oveis Gharan and Luca Trevisan. Partitioning into expanders. In *SODA*, pages 1256–1266, 2014.
- [PSZ14] Richard Peng, He Sun, and Luca Zanetti. Partitioning well-clustered graphs with k-means and heat kernel. *CoRR*, abs/1411.2021, 2014.
- [SJ89] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Inf. Comput.*, 82(1):93–133, 1989.
- [ST96] Daniel A. Spielman and Shang-Hua Teng. Disk packings and planar separators. In *Symposium on Computational Geometry*, pages 349–358, 1996.
- [Tan11] Mamoru Tanaka. Multi-way expansion constants and partitions of a graph. *arXiv.org*, December 2011.
- [Tao12] Terence Tao. *Topics in Random Matrix Theory*. Graduate studies in mathematics. American Mathematical Soc., 2012.
- [VL07] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

## A Proof of Lemma 3.2

*Proof.* For any  $i \leq k$ ,

$$\frac{\xi_i \mathcal{L}_G \xi_i^T}{\xi_i \xi_i^T} = \frac{\mathbf{x}_i D^{1/2} \mathcal{L}_G D^{1/2} \mathbf{x}_i^T}{\mathbf{x}_i D \mathbf{x}_i^T} = \frac{\sum_{(u,v) \in E} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2}{\sum_{u \in V} \deg(u) \mathbf{x}_i^2(u)} = \lambda_i \leq \lambda_k .$$

This further gives that

$$\sum_{(u,v) \in E} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 \leq \lambda_k, \quad (1)$$

since  $\sum_{u \in V} \deg(u) \mathbf{x}_i^2(u) = \sum_{u \in V} \xi_i^2(u) = 1$ .

Let us recall a known result (see, e.g., [Chu97, (1.5), p. 5]) that for any weighted graph  $H = (V_H, E_H)$ ,

$$\lambda_2(H) = \text{vol}_H(V_H) \cdot \min_f \left\{ \frac{2 \cdot \sum_{(u,v) \in E_H} (f(u) - f(v))^2}{\sum_{u,v \in V_H} (f(u) - f(v))^2 \deg_H(u) \deg_H(v)} \right\}, \quad (2)$$

where  $\lambda_2(H)$  denotes the second smallest eigenvalue of the normalized Laplacian of  $H$ .

Let us consider the induced subgraph  $H := G[C]$  on  $C$ . Since  $\varphi(H) \geq \varphi_{\text{in}}$ , Cheeger's inequality yields  $\lambda_2(H) \geq \frac{\varphi_{\text{in}}^2}{2}$ . Therefore, if we apply this bound to inequality (2), then,

$$\text{vol}_H(V_H) \cdot \frac{2 \cdot \sum_{(u,v) \in E_H} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2}{\sum_{u,v \in V_H} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 \deg_H(u) \deg_H(v)} \geq \lambda_2(H) \geq \frac{\varphi_{\text{in}}^2}{2} .$$

Combining this with the fact that  $\sum_{(u,v) \in E_H} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 \leq \sum_{(u,v) \in E_G} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 \leq \lambda_k$ , where the last inequality follows from inequality (1), we have that

$$\sum_{u,v \in V_H} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 \deg_H(u) \deg_H(v) \leq \frac{4\lambda_k \text{vol}_H(V_H)}{\varphi_{\text{in}}^2} .$$

Next, since  $\varphi(H) \geq \varphi_{\text{in}} > 0$  implies that  $\deg_H(u) \geq 1$  for any  $u \in V_H$ , using the bound above we obtain:

$$\sum_{u,v \in V_H} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 \leq \sum_{u,v \in V_H} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 \deg_H(u) \deg_H(v) \leq \frac{4\lambda_k \text{vol}_H(V_H)}{\varphi_{\text{in}}^2} \leq \frac{4\lambda_k \text{vol}(C)}{\varphi_{\text{in}}^2} .$$

□

## Acknowledgment

The authors wish to thank James R. Lee for bringing to their attention a result from the latest version of [LOT12]. This work was partially supported by the NSF grants CCF 1318595 and CCF 1423230.

---

<sup>2</sup>We remark that in [Chu97], the summation in the denominator is over all unordered pairs of vertices, while in our context, the summation is over all possible  $|V_H|^2$  vertex pairs. Therefore, a multiplicative factor 2 appears in the numerator in Equation (2) compared to the form in [Chu97, (1.5), p. 5].